DOCUMENT RESUME

ED 111 871 TM 004 839

AUTHOR Poole, Richard L.
TITLE Student Victimization and the Formulation of Test Construction Criteria.
PUB DATE [Apr 75]
NOTE 18p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, D.C., March 31-April 2, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Elementary Secondary Education; *Evaluation Criteria; Grading; Higher Education; *Student Attitudes; *Student Testing; *Test Construction; *Testing Problems; Test Interpretation; Test Reliability; Test Selection; Test Validity

ABSTRACT

This study surveyed 53 master's level students in two tests and measurement classes to determine if and in what fashion they had been victimized by testing and evaluation. The purpose of this activity was to acquire personal data which could be used to both dramatize and formulate criteria for the construction of classroom tests. A majority of students felt they had been victimized, and their experiences were at all educational levels--elementary and high school; undergraduate and graduate school; in the military and in higher education. These experiences illustrated procedural infringements related to validity, reliability, interpretation, and administration of tests. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED111871

STUDENT VICTIMIZATION AND THE FORMULATION OF TEST CONSTRUCTION CRITERIA

ABSTRACT

This study surveyed 53 master's level students in two tests and measurment classes to determine if and in what fashion they had been victimized by testing and evaluation. The purpose of this activity was to acquire personal data which could be used to both dramatize and formulate criteria for the construction of classroom tests. A majority of students felt they had been victimized, and their experiences were at all educational levels -- elementary and high school; undergraduate and graduate school; in the military and in higher education. These experiences illustrated procedural infringements related to validity, reliability, interpretation, and administration of tests.

TM004 839

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

STUDENT VICTIMIZATION AND THE FORMULATION OF TEST CONSTRUCTION CRITERIA[1]

Richard L. Poole

The National College of Education

Although the criteria for the selection and development of tests are well known and documented (APA, AERA, NCME, 1974) instructional activities devoted to their instillment in students are neither well known nor documented (Mayo, 1967). Indeed most tests and measurement textbooks (Ahmann and Glock 1971, Ebel 1965, and Gronlund 1971) treat this topic, but such treatment appears to be insufficient to induce students or inservice teachers to employ them in their work.

Hence, this study was designed to establish an instructional activity which nurtures teachers to employ test criteria in the construction of their classroom examinations. More specifically, the primary objective was to describe and illustrate an instructional activity for measurement teachers which (a) humanizes and personalizes for the student the identification and formulation

---

1. Paper presented at the annual meeting of the National Council on Measurement In Education, Washington, D.C., April 1975.

Of test construction criteria, (b) dimensionalizes from the students' experience the criteria for the construction of classroom tests, (c) provides the data to inductively formulate the test construction criteria of validity, reliability, interpretability, and administrability. The secondary objective of the study was to show how additional student experiences and data could be acquired for the introduction, discussion, and treatment of many topics germane to tests and measurements.

Method

Two graduate (master's level) tests and measurement classes at a large Eastern state university were asked as an assignment to respond to the following statement made by their instructor:

> The notion has been conveyed to me that students during their educational careers sometime feel as though they have been victimized by testing and evaluation. For next time, would you please tell me in writing if and how you have been victimized by testing and evaluation.

The classes were composed of 27 men and 26 women who were, for the most part, inservice teachers. The number of incidents cited per student was from one to nine with the total number of statements produced being 129. These statements attested to poorly constructed and administered classroom testing or evaluation instruments, misused standardized instruments, and improper grad-

ing policies or procedures. In addition, these statements not only addressed the type of instrument employed-- standardized (STD) or teacher-made (TM), but also the item format employed-- supply (SP) or select (SE). Moreover, despite that fact that the respondents were graduate students their statements indicated that infractions were sustained at all educational levels-- the elementary (ELM) and high (HGH) school (SCH), the undergraduate (UNG) and (GRD) school, as well as in higher education (HED) and the military (MIL). Some of the replies did not indicate where they were sustained and were classified as unspecified (UNS).

It was necessary to increase the areas of infraction, because some of the situations specified on the student replies did not reference the criteria associated with the selection or construction of tests. Some replies were of a miscellaneous or poor testing nature and were accordingly given that label as a descriptor for classification purposes. Other replies were of a social nature and featured either grading, placement and award issues, or testing as punishing, as putting a person down, or as otherwise dehumanizing. In a manner similar to the first case, the labels assigned to these two groups were respectively, "Grading, placement, awards," and "Social insensitivity."

Inasmuch as the replies not only indicated the kind of infraction and where in the student's experience he had sustained the situation, but also the style of the instrument and item

format associated with it, it was then possible to classify replies employing all four groupings. Accordingly instrument style was nested within the kind of infraction, and the item type was nested within where the situation was sustained by the student. Methodologically, the student replies were analyzed, tallied and converted to percentages within a nested two-way classification system.

Results and Discussion

The results of the analysis and classification of the student replies are included in Table 1. Examination of the data suggests that at least two themes were present during the testing and evaluation situations experienced by the 53 surveyed students. One theme related to what might be called technical inadequacies of the tests and evaluation instruments employed, and the other to the human or inhuman characteristics manifested by the teacher during of following the testing or evaluation situation.

In regard to the technical matters the student replies illustrated violations or infringements germane to each test construction or selection criterion. The two test criteria receiving the highest percentage of infractions were those of validity (23.2%) and administrability (21.7%) with the criteria of reliability (7.8%) and interpretability (2.3%) receiving the fewest.

In terms of test type, and for all criteria the teacher-made instrument regardless of its educational level was the

source of three times the amount of criticism as the standardized instrument.

Infractions were sustained at all levels of the students' educational careers, from the elementary school through graduate school, and even in higher education and the military. Twenty-four percent of the replies did not indicate where the situation occurred, but it should be noted that there was an increasing percentage of replies for the elementary school (13%), senior high school (22%) and the undergraduate school (28%). Recency might account for a portion of this result, but for the students surveyed it might also be concluded that long-term memory was operating.

The second theme which appeared in the victimization accounts written by these students, highlighted the human element, or more correctly as stated by them, the inhuman element. According to the student accounts nearly 21% of their teachers added insult to injury by both making derogatory remarks and by belittling them in front of their classmates. The percentage of these indignities was greatest at the high school, and this rate was two and a half times as great as that at the elementary or undergraduate school. In light of the personal sensitivity of most teenagers perhaps this result is not unanticipated, but for the students concerned it apparently did not contribute to their "self concept" or future achievement.

If one expands the test criteria of administrability to refer to the teacher's demeanor during and following the testing or evaluation situation, then the categories of administrability and social insensitivity could be collapsed together. The result of this for the group surveyed would be that the primary area of infraction would no longer be the highly technical criteria of validity, but the less technical and more practical criteria of test administrability. And the conclusion, would be that about half (43%) of the noted testing and evaluation infractions could be eliminated by communicating to teachers the need for being courteous and thoughtful during the administration of classroom tests and when it is being reviewed.

Another result was that six of the 26 women surveyed, or five percent of the group, reported that as far as they could recall, they had not been victimized by testing or evaluation.

Further inspection of Table 1 also shows that 17% of the student replies referenced the issue of grading, placement, or awards, and that the incidences of infraction when comparing teacher-made to standardized instruments is nearly three to one. As recalled by the students, the specific nature of some of these infractions is shown in Table 2. Additionally, Table 2 also contains other unedited student replies for each of the categories. From the student's perspective, the abusiveness of these replies is evident, and unknown from the teacher's, yet the point is to be made that greater understanding and improved use of measurement techniques in education must include the human element.

Writing in the final report of a project entitled "Pre-Service Preparation of Teachers in Educational Measurement," Mayo (1967) indicated that teachers need to acquire competency in the construction and evaluation of classroom tests. He added that "measurement teachers should contrive more ingenious ways to demonstrate the ultimate usefulness of certain competences as they are being learned, "and that test and measurement courses could be improved by developing more meaningful presentations of material.

The activity described here permits the student to glean a perception of test construction and selection criteria which apparently has not been made by the usual textbook-lecture approach. The perception is drawn from the student's own experience and synergistically enlarged from the experiences of his peers. The extent of the personal meaning achieved from this activity appears in the self criticism which students make following the construction and analysis of their own teacher-made achievement tests (for a class they are teaching).

It is unclear to the investigator whether this activity could be called ingenious, but its meaningfulness to the student is certainly clear. As one student wrote at the end of his reply:

> "A teacher should recall any personal detriment suffered due to testing inconsistencies and analyze them for guideline to be used in the construction of more meaningful tests."

| Where Sustained | | Area of Infraction | | |

TABLE 2

Illustrative Student Replies by Category

Category	Reply
Validity	
1.	During a psychology course recently I was "victimized" as were the other students by the evaluation method used. The professor lectured to the class on the material which probably was most interesting to him or which he could add his own comical-sarcastic adlibs. His classes were enjoyable because he did entertain well. When it came to testing, however, he went completely to the book and chose the smallest most factual material that he could find for at least 60% of the test. There was little correspondence between class material and testing material.
2.	My first response was - "I've never been victimized by a test. I always did well, and therefore have nothing to complain about.... "But the more I thought about the question, the more I realized that I wasn't victimized by the tests themselves because I knew how to play the game. That is, I knew how to memorize and was an absolute ace at recall. Now, as I have thought through the question and its implications, I feel a cold anger not only in regard to classroom testing, but to the philosophy and the objectives behind the tests. Evaluation

consisted of testing ability at recall and memorization -- it did not test ability to think, to solve problems, or to apply knowledge in a creative way.

3. Styles come and go in education and so the "progressive" principal of my public school in city X in the 1920's decided to give intelligence tests to all the elementary pupils. Because I did well with good visual memory and the inner motivation to do well, they decided to skip me several half grades in the lower grades. Both of my parents had been teachers and saw nothing wrong in this procedure since I was capable of assimilating material and was serious about school marks, etc. However, I was the slow maturing type and this advancement in school threw me sadly out of pace with my fellow students. My social development got stuck, I became a "loner" with the exception of one friend with whom I did many things outside of school before graduation in eighth grade. I had trouble joining in with my peers all during high school and even an extra post graduate year in high school to get me to the age of 17 before entrance into college did not mend the gap socially.

Administrability

4. I don't remember the exact times, but the following situation has happened to me more than once. This is the time when the teacher is "kind" enough to tell you what to study, throws

you a curve, and you do not have any idea what the test or final is about.

5. On the high school level, most tests seemed pretty fair except that their directions were often vague. Most students just assumed a set way of answering questions and then answered them without paying too much attention to the directions.

6. As a student in grade school, I was "victimized" by classroom testing as follows: I was a consistently good student, but one day the class got back some arithmetic tests and my grade was F. It seemed unlikely to my parents so we checked the problems and many marked wrong were seemingly correct. Reviewing it with the teacher, it was discovered that I had copied numbers incorrectly from the board. (As it turns out we discovered I was quite nearsighted.) The testing was unfair since problems written on the blackboard in a large room may not be equally clear to all students.

Social Insensitivity

7. First of all, remembering back, my first experience of this was being threatened by a teacher that we would have to take a series of tests unless we started to behave outselves.

8. . . . a list of English words were given and we were asked to make sentences with them. I remember mistaking the word "shrivel" for "shiver" and made the sentence: "Don't shrivel with coal (for cold) wear your sweater." The teacher copied

the sentence with my name on it and pasted it on the bulletin board. The other students made fun of me and were very cruel apparently without meaning to be. But after that incident, I never wanted to write anything for this particular teacher.

9. . . . When I inquired into why I received a low grade, the response was "Anyone who receives an "E" (failing grade) on any exam, never deserves a grade of "B" in my course!

Reliability

10. When I was in the 9th grade my history teacher asked a true-false question I felt was ambigious. I was a good student and the incorrect answer did not effect my grade, however, I was annoyed because it didn't tap my knowledge of the subject but rather my interpretation of what the word important meant as compared with the teacher's interpretation of the word.

In this case, neither my understanding or factual knowledge of the subject were tapped by the question. This happened a dozen years ago, yet, I can still remember the question and how and why I answered it "incorrectly".

11. . . . Other times of being victimized by a test occurred when some factor (external or internal) has distracted me to such an extent that I was not able to perform as expected. Such distractions have taken the form of noise, being slightly ill, etc. This distraction has also been of the form of the make-up of the test. This happened when the test activity

was unfamiliar, that is not of the form that was used in the instruction.

Interpretability

12. . . . The test wasted the person's time in taking it, was unfair to the student who worked hard to get a good grade, and gave no information to me as to whether I really learned the material required in the course.

13. . . . no feedback.

Grading

14. I received a "99" on my 8th grade English final exam. My teacher said that I answered all the questions correctly but she had to deduct one point from my composition because she didn't feel that anyone deserved a perfect grade in English:

15. My example deals more with being victimized as a result of evaluational procedures rather than actual testing, although testing did play a role. During my work as an undergraduate, I was required to take a course in Statistical Psychology in order to take a course that I wanted to take in abnormal psychology. The course was designed for psy majors and was in some sense used as a basis for weeding out the unwanted excess. Being a math major, I felt that I would have little trouble.

The grade was determined by evaluating performances on three hourly examinations, one final and a lab project. On two of the hourly exams I scored A, on the third C. Seeing that we all

strive for that "A", I sought advice from the instructor. I was told that all I needed to do was to score high on the final and get an A on the project (the project being more important). My project was graded A, and my final A, but somehow I was given a B. When I returned to question the grade, I found that the instructor had left for the summer. An assistant offered an explanation based upon "cut-off" points. You see, they could only award a certain number of A's.

16. . . . One practice that is very bad is that of reserving the high grades for majors.

17. . . . I was then flabbergasted when she (the teacher) nearly failed me at the final test in which I had put up my best performance. When I went to find out why I was graded so low she said I had been cheerful in class and she did not like cheerful people. This instructor in her dealings with me at this instance let her emotions dictate for her how to evaluate my work.

Poor Testing

18. I'm not exactly sure exactly "how" I was victimized through testing and evaluation procedures in my past.

That I have been victimized, there is no doubt, for in reading the yellow booklet (

IMproving the Classroom Test, A Manual of Test Construction; Procedures for the Classroom Teacher; The State Education DEpartment, Bureau of Examinations and Testing, Albany, 1958.

on test construction I find that my own tests contain many faults when analyzed according to the checklist in the rear of the book. I was never taught how to construct tests, therefore I suppose I constructed them partially by "common sense" and by the influence of tests I took in the past (conditioning, if you like).

References

Ahmann, J.S. and Glock, M.D., Evaluating Pupil Growth. (4th Ed.) Boston: Allyn and Bacon, 1971.

American Psychological Association, American Educational Research Associate, National Council on Measurement in Eudcation, Standards for Educational and Psychological Tests. Washington: American Psychological Association, 1974.

Ebel, R.L., Measuring Educational Achievement. Englewood Cliffs: Prentice Hall, 1965.

Gronlund, N.E., Measurement and Evaluation in Teaching. (2nd Ed.) New York: Macmillan, 1971.

Mayo, S.T. Pre-Service Preparation of Teachers In Educational Measurement. Chicago: Loyola University, December 1967.

Project No 5-0807, Contract No. OE 4-10-011, U.S. Department of Health, Education, and Welfare.